

# 담화표지의 음성언어적 특성과 음성합성 시스템에서의 활용

이호준      박종철  
한국과학기술원 전자전산학과 전산학전공  
{hojoon, park}@nlp.kaist.ac.kr

## Characteristics of Spoken Discourse Markers and their Application to Speech Synthesis Systems

Ho-Joon Lee      Jong C. Park  
Computer Science division, EECS department, KAIST

### 요 약

음성은 컴퓨터로 대변되는 기계와 사람 그리고 기계를 매개로 한 사람과 사람의 상호작용에서 가장 쉽고 직관적인 인터페이스로 널리 활용되고 있다. 인간에게 음성정보를 제공하는 음성합성 분야에서는 합성결과의 자연스러움과 인식성이 시스템의 주요 평가요소로 활용되고 있는데 이러한 자연스러움과 인식성은 합성결과의 정확성뿐만 아니라 발화환경이나 발화자의 발화특징 혹은 감정상태 등에 의해 많은 영향을 받게 된다. 담화표지는 문장의 명제 내용에는 직접 관여하지 않으면서 화자의 발화 의도나 심리적 태도를 전달하는 구성 요소를 말하는데 본 논문에서는 담화표지가 포함된 대화 음성 데이터를 수집하여 담화표지의 음성언어적인 특징을 분석하고 분석된 결과를 음성합성 시스템에 활용하는 표현 방식에 대해 논의한다.

### 1. 서론

인간과 인간 사이에서 가장 기본적인 의사소통방법 중 하나인 음성은 기술의 발달과 함께 컴퓨터로 대변되는 기계와 사람 그리고 기계를 매개로 한 사람과 사람의 상호작용에 가장 쉽고 직관적인 인터페이스로 활용되고 있다. 인간과 기계 사이의 음성 인터페이스는 주로 인간의 음성을 인식하는 음성인식 분야와 인간에게 음성정보를 제공하는 음성합성 분야로 나눌 수 있는데 음성합성 분야에서는 합성결과의 자연스러움(naturalness)과 인식성(intelligibility)이 음성합성 시스템의 주요 평가요소로 활용되고 있다.

이러한 자연스러움과 인식성은 합성결과의 정확성(correctness)과는 다르게 주관적인 성격을 지니기 때문에 발화환경이나 발화자의 발화특징 혹은 감정상태 등에 의해 많은 영향을 받게 된다. 이 중 감정상태에 따른 발화문의 음운적 특징에 대한 연구는 국내외에서 활발하게 진행되고 있는데 주로 음의 높낮이, 길이 및 세기의 변화를 기반으로 음운론적인 연구가 이루어지고 있다. 개개인의 주관적인 발화 특징은 발화문의 음운론적인 요소에서만 아니라 발화문장 자체에서도 나타나는데 그 중 문장의 명제 내용에는 직접 관여하지 않으면서 화자의 발화 의도나

심리적 태도를 전달하는 문장의 구성 요소를 담화표지라고 한다[1].

한국어의 담화표지에 관한 연구는 그 기능과 의미의 측면에서 감탄사나 군말, 간투사, 부사, 불변화사, 화용표지[2] 등으로 다양하게 이루어져 왔는데 그 결과 담화표지의 정의에 대해서는 서로 유사성을 지니고 있지만 그 범위나 인식방법에 대해서는 다양한 의견이 있으며 이러한 문제는 영어권에서도 동일하게 나타나고 있다[3]. 그러나 한국어의 담화표지에 관한 연구는 주로 구어를 문자화한 자료를 기초로 하여왔기 때문에[4] 담화표지의 유형 및 의미 분석이나 문법화 과정에 집중되었고 구어에서 담화표지가 지닌 음성언어의 특징을 파악하기는 어렵다는 문제점을 지니고 있다[2][5][6].

따라서 본 논문에서는 적절한 형태의 담화표지가 표현된 음성 데이터를 수집하여 담화표지의 음성언어적인 특징을 분석하고, 음성합성 시스템에서 활용하기 위한 담화표지의 음성언어적 표현방식에 대해 논의하고자 한다.

이어지는 2 절에서 담화표지에 대한 선행연구의 흐름 및 특징을 살펴본 뒤 3 절에서 담화표지가 포함된 음성 데이터의 수집 및 음운자질 분석에 대해 논의하고 4 절에서 음성합성 시스템으로의 활용방안을 보인다.

## 2. 선행연구

국어 화용표지의 연구[2]에 따르면 한국어의 담화표지를 화행이나 문장 또는 그보다 작은 언어적 단위를 묶어주면서 담화의 선후 의존성(sequential dependence)을 표현하는 기능을 갖기도 하는, 담화의 의미적 결속을 위한 선택적 장치라고 정의하고 있는데, 앞서 논의한 바와 같이 한국어의 담화표지에 관한 연구는 주로 문자화된 담화표지의 유형과 문법화 과정에 중점을 두어 진행되어 왔다.

한국어 담화표지 분석[7]에 관한 연구에서는 담화표지의 기능을 부름표지, 시발표지, 전환표지, 결말표지로 분류하였고, 국어 담화표지 ‘인자’에 대한 연구[1]에서는 담화표지를 어휘적 담화표지와 비어휘적 담화표지로 분류하여 비어휘적 담화표지로 지시어 담화표지와 감탄사 담화표지, 초분절음 담화표지를 제시하였다. 국어 담화표지의 유형과 담화표지되기[8]에서는 감탄사, 부사어, 지시어, 운소(억양, 쉼, 강세)등을 본디 담화표지로 구분하고 내용어나 기능어에서 담화표지로 변화된 형태를 전성 담화표지로 구분하여 전성 담화표지를 문법화의 틀에서 설명하고자 하였다. 담화표지되기와 문법화[9]의 연구에서는 이러한 전성 담화표지를 명사, 대명사, 관형사, 부사, 명사구, 동사구로 더욱 세분화하여 설명하고 있으며 이러한 연구와 함께 담화표지 ‘뉘’[10]나 ‘글쎄’[11] 등 개별적인 어휘에 대한 담화 기능 및 의미 분석도 함께 이루어져 왔다.

한편 성별에 따른 국어 담화 표지 사용 모습[4]에 관한 연구에서는 전국 중고등학교 이야기대회에서 수집한 발화문을 문자화하여 담화표지의 사용빈도 및 그 유형 등을 발화자의 성별과 연관지어 분석하였는데, 이러한 분석 자료를 활용하면 발화자의 연령대나 거주지역 등과의 연관성도 찾을 수 있을 것으로 예상된다.

영어권에서 담화표지(discourse marker)에 대한 연구는 언어학적인 관점에서뿐만 아니라 전산학적인 관점에서도 활발히 이루어지고 있는데, 담화표지의 자동인식에 관한 연구[3]에서는 담화표지의 음성언어적인 특징(담화표지 전후의 휴지길이 및 담화표지의 발화길이)이 담화표지의 인식률에는 큰 영향을 주지 못하였지만 이러한 음성언어적 특성의 활용 가능성 및 음의 높낮이, 발화속도 등 추가적인 음운 자질의 활용에 대해서 논의가 있었다. 또한 억양경계(intonational boundaries)와 말 바꾸기(speech repair), 담화표지와 상관계에 대한 연구[12]에서는 각 항목에 일종의 품사를 부여한 언어 모델을 제시하고 확률 모델을 통해 억양경계, 말 바꾸기, 담화표지를 인식하는 방법을 제안하였다.

이상에서 살펴본 담화표지에 관한 대부분의 연구는 발화문을 문자화하여 담화표지의 유형을 분류하고 담화표지 현상을 언어학적인 틀에서 설명하고자

하였으며 전산학적인 관점에서도 주로 음성인식 분야에서 인식결과를 높이는 방향으로 연구가 진행되었는데 담화표지의 음성언어적 분석은 자연스러운 합성결과를 위한 음성합성 시스템 등에서도 다양하게 활용이 가능할 것으로 예상된다.

## 3. 음성 데이터 수집 및 분석

2절에서 살펴본 선행연구 중 다수는 수집된 음성 데이터를 문자화하여 구어의 음운적 특징을 쉽게 확인할 수 없을 뿐만 아니라 자유로운 발화의 형태가 아닌 미리 쓰여진 문장을 발화하는 상황에서 음성 데이터를 수집하여 실제 담화표지가 가지는 구어환경에서의 특징을 나타내는데 어려움이 있다. 그로 인하여, 수집된 담화표지의 형태가 주로 발화자의 지속적인 발화습관이나 ‘에’나 ‘음’ 등과 같은 채움말(filler)의 형태로 표현되는 것을 확인할 수 있다.

본 논문에서는 쓰여진 문장을 읽을 때 많이 나타날 수 있는 반복적인 발화자의 발화습관이나 채움말 등의 표현 이외에, 자유로운 대화에서 담화표지가 가지는 음성언어적 특징을 분석하고 이를 음성합성 시스템을 통해 표현하는 방법에 대해 논의하고자 한다.

### 3.1. 음성 데이터 수집

담화표지의 화용적, 담화적 기능을 분석하기 위해서는 음성언어 형태의 발화문 외에도 발화자의 표정, 몸짓, 주변 인물 등의 발화 환경을 담은 영상자료가 활용될 수 있기 때문에 본 논문에서는 토크쇼 형태의 TV 프로그램 분석 도메인으로 선정하였다. TV 드라마의 경우에는 다양한 발화환경을 가지는 장점이 있지만 미리 쓰여진 대본을 바탕으로 발화가 이루어진다는 점에서 뉴스 도메인에서와 같이 담화표지의 표현이 작위적일 수 있다는 문제점이 있다. 본 논문에서는 담화표지의 음운 자질을 분석하기 위해 MBC TV의 ‘무릎팍도사’ 5회분의 자료를 분석하여 담화표지가 포함된 문장 22개를 추출하였다. 2절의 선행연구에서 살펴본 바와 같이 한국어 담화표지의 범위에 대해 다양한 해석이 가능한데 본 논문에서는 발화문장에서 특정한 의미 없이 부가적인 형태로 사용된 ‘그, 뉘, 막’ 등의 세가지 표현에 국한하여 데이터를 수집하였다.

토크쇼 한 회의 분량이 40~50분임을 감안할 때 추출된 담화표지의 양이 매우 적었는데, 특정 인물의 경우에는 발화를 하는 도중에 담화표지를 거의 사용하지 않는 등의 흥미로운 특징도 발견되었다.

### 3.2. 담화표지의 음의 길이 분석

본 논문에서는 수집된 영상데이터에서 음성 데이터만을 별도의 압축과정 없이 추출하여 Praat을 이용하여 담화표지의 음운자질을 분석하였다. 추출한 담화표지의 음운 자질은 담화표지의 발화길이 및 담화표지의 앞뒤에 나타나는 휴지의 길이와 문장 내에서의 음의 높낮이 변화, 그리고 음의 세기의 변화이다.

한 명의 발화자로부터 수집한 담화표지 ‘그’에 대한 음의 발화길이는 아래의 표 1과 같다.

표 1 발화자 1의 담화표지 ‘그’의 음의 길이

문장	길이 (msec)	앞 휴지 길이 (msec)	뒤 휴지 길이 (msec)
본부에서 그 정식 그 비타민 프로그램에서	12.5	없음	없음
본부에서 그 정식 그 비타민 프로그램에서	31.9	없음	22.6
입장에서는, 그 김구라씨	14.1	75.1	없음
아무튼 그 그때	54.15	없음	167.6

표 1에서 15msec 내외로 발화되는 ‘그’의 경우 추가적인 휴지가 담화표지의 전후에 나타나지 않지만 30msec 이상으로 길게 발화되는 ‘그’의 경우에는 담화표지에 바로 이어서 긴 휴지가 나타나는 것을 확인할 수 있다. 일반적인 길이로 발음되는 ‘그’의 경우는 바로 이어지는 어휘를 강조하는 효과 정도로 사용되고 있지만 길게 발음되는 ‘그’의 경우에는 바로 이어지는 휴지와 함께 채움말의 형태로도 사용되는 것을 확인할 수 있다. 동일한 담화표지 ‘그’에 대해서 다른 발화자의 발화 데이터를 정리하면 표 2와 같다.

표 2 발화자 2의 담화표지 ‘그’의 음의 길이

문장	길이 (msec)	앞 휴지 길이 (msec)	뒤 휴지 길이 (msec)
굉장히 그 거친 친구들이	8.9	없음	없음
제 방송을 예전에 그 듣고 이랬던 친구들이	7.8	없음	없음
원래 그 불평 불만 이라던지 이런게 굉장히	5.6	없음	없음
어쨌든 뭐 제가 그 그 당시 뭐 거칠게 방송할	13.1	없음	없음
저는 그 인터넷 이런 거 들어가서요	50.9	없음	52.4

표 1에서의 결과와 마찬가지로 담화표지 ‘그’가 짧은 음의 길이를 가지는 형태로 사용될 때에는 담화표지의 앞, 뒤에 추가적인 휴지가 나타나지 않지만, 일반적인 발화 음절보다 길게 발화되는 경우에는 담화표지에 바로 이어서 긴 휴지가 함께 나타나며 채움말의 형태로도 사용되는 것을 확인할 수 있었다. 아래 표 3을 보면, 동일한 발화자가 사용한 담화표지 ‘뭐’의 분석결과 역시 이와 유사한 형태를 보인다.

표 3 발화자 2의 담화표지 ‘뭐’의 음의 길이

문장	길이 (msec)	앞 휴지 길이 (msec)	뒤 휴지 길이 (msec)
그것도 이제 뭐 궁극적인 목표일 수도 있습니다만	15.3	없음	없음
제가 의도적으로 뭐 이 사람들을	17.4	없음	없음
그 당시 뭐 거칠게 방송할 때	26.7	없음	11.6
그 당시에 이제 뭐 제가	28.3	없음	없음
저의 집사람 입장에서는 제가 뭐 나름대로	9.4	없음	없음
거기 있는 뭐 수치를 좀 외우려고 하고	13.3	없음	없음

담화표지 ‘그’나 ‘뭐’에 비해서 많이 나타나지는 않았지만 담화표지 ‘막’이 사용된 예를 살펴보면 표 4와 같다.

표 4 발화자 3과 4의 담화표지 ‘막’의 음의 길이

문장	길이 (msec)	앞 휴지 길이 (msec)	뒤 휴지 길이 (msec)
저희 고등학교에 막 플래카드 붙고	15.23	없음	38.57
이렇게 막 사람들이 작아	6.4	없음	없음

일반적으로 담화표지는 담화표지의 앞뒤에 별도의 휴지 없이 평균적인 발화 길이보다 짧게 발화되는 특징을 보였고 담화표지가 길게 발화되는 경우에는 발화길이와 비슷한 정도의 길이로 휴지가 이어져 나타나는 것을 확인할 수 있었다. 전자의 경우는 담화표지가 담화표지 전후의 단어의 의미를 강조하는 효과로 사용된 것으로 보여지며 후자의 경우는 긴 발화시간과 바로 이어지는 긴 휴지길이를 채움말의 효과를 함께 보이는 것으로 해석할 수 있다.

그림 1의 붉은 경계는 표2의 마지막 문장 ‘저는 그 인터넷 이런 거 들어가서요’에서 담화표지 ‘그’와 이어져 나타나는 휴지의 길이를 Praat에서 분석한 결과를 보여주고 있으며 그림 2의 붉은 경계는 표4의 첫번째 문장 ‘저희 고등학교에 막 플래카드 붙고’에서 담화표지 ‘막’의 발화길이와 이어져 나타나는 휴지 길이를 보여준다.

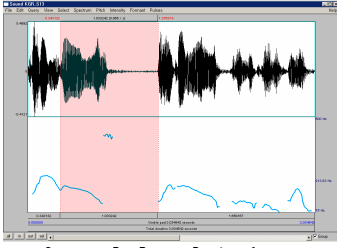


그림 1 담화표지 ‘그’

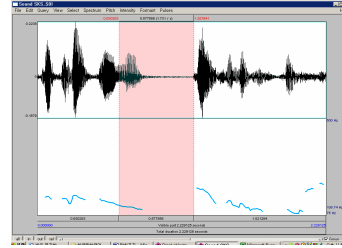


그림 2 담화표지 ‘막’

### 3.3. 담화표지의 음의 높낮이 분석

음의 높낮이(pitch)는 발화문의 운율정보 중에서 화자의 의도나 감정상태 등의 정보를 표현하는데 가장 중요한 요소로 알려져 있다. 본 절에서는 음성언어로 표현된 담화표지의 음의 높낮이 변화를 전체 발화문의 음의 높낮이와 비교하고 음의 높낮이 곡선의 형태를 분석하여 음성합성 시스템에 활용하는 방안을 제시한다.

표 5는 표1에서 살펴본 문장의 평균 음의 높낮이와 문장에서 표현된 담화표지의 평균 음의 높낮이, 그리고 음의 높낮이 곡선의 분석 결과를 보여준다.

표 5 발화자 1의 담화표지 ‘그’의 음의 높낮이

문장	문장 평균 높낮이 (Hz)	담화표지 평균 높낮이 (Hz)	곡선 형태
본부에서 그 정식 그 비타민 프로그램에서	169.74	205.81	단순 감소
본부에서 그 정식 그 비타민 프로그램에서	169.74	209.95	단순 감소
그 김구라씨 인터넷 방송이	186.97	172.18	단순 감소
아무튼 그 그때 당시에	212.26	206.63	단순 감소

표 6 발화자 2의 담화표지 ‘뭐’의 음의 높낮이

문장	문장 평균 높낮이 (Hz)	담화표지 평균 높낮이 (Hz)	곡선 형태
그것도 이제 뭐 궁극적인 목표일 수도	174.01	186.38	단순 감소
어쨌든 간에 제가 의도적으로 뭐 이 사람들을	177.86	186.64	단순 감소
그 당시 뭐 거칠게 방송할 때 본의 아니게	102.6	114.57	단순 감소
그 당시에 이제 뭐 제가	174.37	154.2	단순 감소
제가 뭐 나름대	113.64	126.02	단순

로			감소
거기 있는 뭐 수치를 좀 의우려고 하고	119.39	96.21	단순 감소

표 6은 표3에서 살펴본 문장의 음의 높낮이 분석결과를 나타내는데 표5와 비교하여 볼 때 담화표지의 평균 음의 높낮이는 전체 문장의 평균 음의 높낮이와 유사한 범위에서 변화하고 있으며 음의 높낮이 변화 형태가 그림 1과 2에서 파란색 점선으로 보인 것과 같이 모두 단순 감소하는 형태인 것을 확인할 수 있다. 이러한 특징은 담화표지가 전형적인 감탄사의 복잡한 음의 높낮이 변화 모습과는 다르게 정적인 형태로 표현되며, 따라서 담화표지가 음의 높낮이 변화에 따른 영향을 거의 받지 않을 것으로 해석할 수 있다. 그러므로 담화표지를 음성합성 시스템에서 활용할 때에는 주변 발화문과의 평탄화 과정(smoothing)을 통해 음의 높낮이 변화를 주변 음절의 수준으로 제한할 필요가 있다.

그림 3의 붉은 표시는 표4의 두번째 문장 ‘이렇게 사람들이 막 작아’의 전체 음의 높낮이 변화에서 담화표지 ‘막’의 위치를 보여주고 있다. 전체적으로 볼 때, 이 문장에서는 음의 높낮이가 매우 다양하게 변화하고 있지만 담화표지 부분을 확대한 그림 4를 살펴보면 담화표지 ‘막’의 경우 음의 높낮이 변화가 거의 없는 것을 확인할 수 있다.

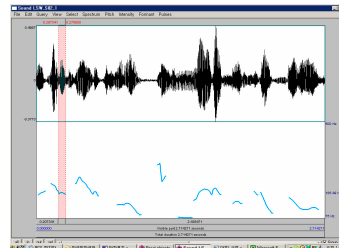


그림 3 담화표지 ‘막’

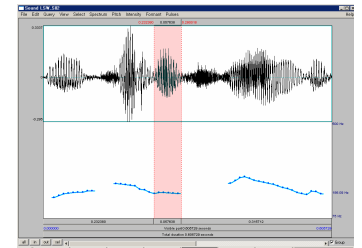


그림 4 세부적 음의 높낮이

### 3.4. 담화표지의 음의 세기 분석

담화표지의 음의 세기를 상대적인 관점에서 분석하기 위해 아래의 표와 같이 담화표지가 포함된 문장의 평균 음의 세기를 측정하고 담화표지의 평균 음의 세기와 차이를 백분율로 표현하였다.

표 7을 살펴보면 담화표지의 평균 발화세기가 문장의 평균 발화세기보다 약 5%정도 더 큰 것을 확인할 수 있는데 이와 같은 특징은 표 8, 표 9, 표 10에서 보이는 바와 같이 본 논문에서 수집한 대부분의 발화문에서 약간의 오차를 보이며 유사하게 나타났다.

표 7 발화자 1의 담화표지 ‘그’의 음의 세기

문장	문장 평균세기 (dB)	담화표지 평균세기 (dB)	차이 (%)
본부에서 그 정식 그 비	74.7	77.6	3.83

타민 프로그램에서			
본부에서 그 정식 그 비타민 프로그램에서	74.7	76.5	2.41
지금 생각해보면 그 연예인 입장에서는, 그 김구라씨 인터넷 방송이 어떻게 와 닿았을까요?	71.8	76.4	6.41
아무튼 그 그때 당시에 피해를 봤던 동료 연예인들에게 사과할 마음이 없다는거죠	74.8	78	4.29

표 8 발화자 2의 담화표지 ‘그’의 음의 세기

문장	문장 평균세기 (dB)	담화표지 평균세기 (dB)	차이 (%)
제 팬들이 굉장히 그 거친 친구들이 많아요	69.8	72.6	4.01
제 방송을 예전에 그 듣고 이랬던 친구들이 굉장히 거친	72.3	77.1	6.61
제가 원래 그 불평 불만이라던지 이렇게 굉장히 많습시다	70.1	73	4.14
어쨌든 뭐 제가 그 그 당시 뭐 거칠게 방송할 때 본의 아니게 제 입에 의해서	71.6	79.1	10.4
저는 그 인터넷 이런 거 들어가서요 게임 안합니다	71.3	75.3	5.58

표 9 발화자 2의 담화표지 ‘뭐’의 음의 세기

문장	문장 평균세기 (dB)	담화표지 평균세기 (dB)	차이 (%)
그것도 이제 뭐 궁극적인 목표일 수도 있습니다	69.8	71.5	2.45
제가 의도적으로 뭐 이 사람들을	71.2	75.5	6.1
그 당시 뭐 거칠게 방송할 때 본의 아니게 제 입에 의해서	71.9	74.3	3.45
그 당시에 이제 뭐 제가 저희 아버님께서도 아프셨어요	71.3	74.7	4.74
제가 뭐 나름대로 월척은 아니어도	72.1	72.8	0.99
거기 있는 뭐 수치를 줘	71.7	75.1	4.68

외우려고 하고			
---------	--	--	--

표 10 발화자 3과 4의 담화표지 ‘막’의 음의 세기

문장	문장 평균세기 (dB)	담화표지 평균세기 (dB)	차이 (%)
저희 고등학교에 막 플래카드 붙고	62.2	61.2	-1.6
이렇게 막 사람들이 작아, 이런 텔 갈 수가 없어요	69	71.3	3.33

위의 표에서 살펴본 바와 같이 담화표지에서의 음의 세기가 문장의 평균적인 음의 세기보다 큰 이유는 본 논문에서 담화표지를 한 음절의 ‘그’, ‘뭐’, ‘막’ 등으로 제한하였기 때문일 수도 있지만, 음의 높낮이, 길이, 세기 등 운율정보가 상호 유기적으로 영향을 준다고 볼 때 3.3절에서 살펴본 담화표지의 단순한 음의 높낮이 변화에 의한 결과라고도 볼 수 있다. 즉 발화문에서 나타나는 담화표지의 음의 높낮이에는 큰 변화가 없기 때문에 음의 세기도 큰 변화 없이 일정한 크기로 표현되어, 복잡한 음의 높낮이를 보이는 전체 발화문장에 비해 담화표지의 평균 발화세기가 크게 분석된 것으로 볼 수 있다.

지금까지는 수집한 음성 데이터에서 담화표지가 가지는 음성언어적 특성을 음의 길이와 휴지의 길이, 음의 높낮이, 음의 세기 등의 운율정보를 바탕으로 분석하였다. 4절에서는 이러한 정보를 이용하여 음성합성 시스템에 활용하는 방법에 대해서 논의한다.

#### 4. 담화표지를 활용한 음성합성 시스템

자연스러운 음성합성을 위해서는 정확한 음성표현을 기반으로 발화환경이나 발화자의 발화특징 혹은 감정상태 등을 함께 표현할 수 있어야 한다. 따라서 구어에서 특징적으로 나타나는 담화표지의 적절한 표현은 음성합성 결과의 자연스러움을 향상시킬 수 있는데, 본 논문에서는 3절에서 확인한 담화표지의 음운특징을 아래와 같이 단순화 하여 SSML(Speech Synthesis Markup Language)을 이용하여 담화표지를 표현하였다.

- 담화표지의 발화속도가 빠르면 담화표지의 전후에 추가적인 휴지가 없다.
- 담화표지의 발화속도가 느리면 담화표지 다음에 동일한 길이의 추가적인 휴지가 있다.
- 담화표지의 음의 높낮이는 담화표지 전후 어절의 평균값을 가지며 하향곡선을 그린다.
- 담화표지의 음의 세기는 담화표지 전후 어절보다 5% 큰 값을 갖는다.

SSML은 웹 환경에서의 음성합성을 위한 XML 기반의 언어로 W3C 컨소시엄에 의해 2004년에 표준화가

이루어졌다. SSML에서는 prosody라는 태그를 사용하여 음의 높낮이, 말의 빠르기, 소리의 크기를 조절할 수 있는데 SSML을 이용하여 느린 형태의 담화표지 ‘뭐’를 표현하면 아래 그림 5와 같다.

```
<?xml version="1.0"?>
<speech version="1.0" xmlns="http://www.w3.org/2001/10/synthesis"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.w3.org/2001/10/synthesis
    http://www.w3.org/TR/speech-synthesis/synthesis.xsd"
  xml:lang="UTF-8">
  <prosody rate="-20%" contour="(70%,-5%) (100%,-20%)" volume="+5%">
  뭐
  </prosody>
  <break time="20ms"/>
</speech>
```

그림 5 담화표지 ‘뭐’의 SSML 표현

그림 5를 살펴보면 담화표지 ‘뭐’에 대해서 prosody 태그를 이용하여 전체의 발화속도를 20% 느리게 하고 음의 높낮이는 5% 감소에서 20% 감소의 형태를 변화를 주며 발화크기를 5% 크게 설정하고, 바로 이어서 20ms의 휴지를 가질 수 있도록 표현하였다.

본 논문에서는 Voiceware의 VoiceText Text-to-Speech 엔진을 이용하여 SSML을 지원하는 웹 기반의 음성합성 시스템으로 합성 결과를 생성하였다. 현재까지는 SSML을 통한 음의 높낮이 곡선(pitch contour)의 표현이 음성합성 시스템에서 완벽하게 지원되지 않기 때문에 사용자가 어절 단위로 음의 높낮이와 빠르기, 세기를 3단계로 조절하여 표현할 수 있게 하였다.

본 연구에서 구현한 음성합성 시스템을 이용하여 표2의 마지막 문장 ‘저는 그 인터넷 이런 거 들어가서요’를 SSML의 형태로 변환하고 음성으로 합성한 결과는 그림 6과 같다.

```
<?xml version="1.0"?>
<speech version="1.0" xmlns="http://www.w3.org/2001/10/synthesis"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.w3.org/2001/10/synthesis
    http://www.w3.org/TR/speech-synthesis/synthesis.xsd"
  xml:lang="euc-KR">
  <prosody pitch="medium" rate="medium" volume="medium">저는</prosody>
  <prosody pitch="medium" rate="slow" volume="loud">그</prosody>
  <prosody pitch="medium" rate="slow" volume="medium"></prosody>
  <prosody pitch="default" rate="default" volume="default">인터넷</prosody>
  <prosody pitch="default" rate="default" volume="default">이런</prosody>
  <prosody pitch="default" rate="default" volume="default">거</prosody>
  <prosody pitch="default" rate="default" volume="default">들어가서요</prosody>
</speech>
```



재생

그림 6 담화표지의 SSML 표현 및 음성표현 생성

제안한 웹 서비스를 통해 합성된 결과를 다시 Praat으로 분석한 결과는 그림 7과 같다. 붉은 색으로 표시된 부분이 담화표지 ‘그’를 나타내는데 초록색 선은 음의 세기 변화를, 파란색 선은 음의 높낮이 변화를 보여준다.

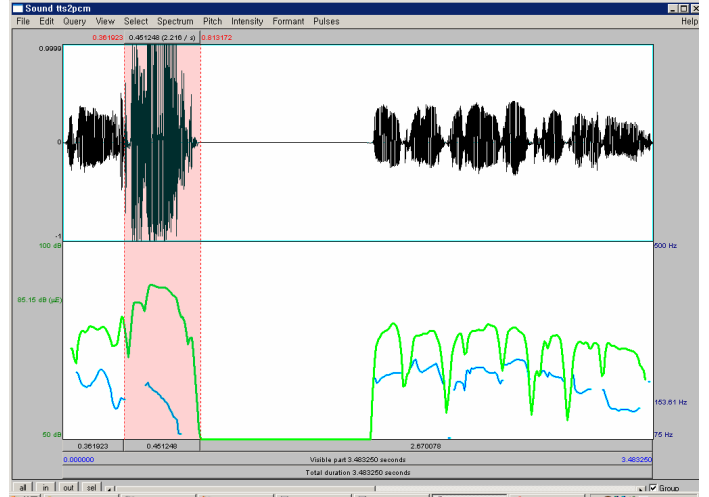


그림 7 합성된 음성결과 분석<sup>1</sup>

그림 7의 분석결과를 원본 발화문의 분석결과인 그림 1과 비교해보면 음의 길이와 휴지의 길이는 예상했던 바와 같이 유사한 모습을 보이지만 음의 높낮이의 경우, 담화표지 ‘그’를 제외하고는 뚜렷한 유사성을 찾기 어렵다. 이는 SSML로 음운정보를 표현할 때 어절단위로 정보가 표현되기 때문에 어절 내에서는 세부적인 음운정보의 조절을 하기 어렵기 때문이다. 그렇지만 담화표지의 경우 음의 높낮이나 음의 세기가 급격하게 변화하지 않는 것으로 확인되었고, 본 논문에서 다루는 담화표지가 한 음절의 형태이기 때문에 SSML을 이용한 표현으로도 원본에 유사한 결과를 만들 수 있었다. 따라서 앞으로 음의 높낮이 곡선(pitch contour) 등의 세부적인 음운 조절방식이 SSML 기반의 음성합성 시스템에 완벽하게 지원된다면 더욱 자연스러운 결과를 생성할 수 있을 것으로 예상된다.

## 5. 결론 및 향후계획

본 논문에서는 담화표지가 포함된 자연스러운 대화의 음성 데이터를 수집하여 담화표지의 음성언어적인 특징을 음의 길이와 휴지, 음의 높낮이, 음의 세기로 세분화하여 분석하고 그 결과를 음성합성 시스템에 활용하는 방안에 대해서 논의하였다. 그 결과 담화표지가 일반적인 감탄사와는 다르게 정적인 형태의 운율구조를 가진다는 점을 확인하였으며, 담화표지의 발화 길이와 이어나오는 휴지의 길이 사이에 밀접한 연관성이 있음을 확인하였다. 또한 이러한 음운특징을 단순화하는 과정을 통해 SSML로 담화표지가 포함된 음성언어를 표현하였고, 음성합성 시스템을 구축하여 실제 합성음을 생성하였다.

SSML의 경우 음성합성 시스템에 종속되지 않고 음운특징을 명시적으로 표시해 줄 수 있다는 장점이 있지만

<sup>1</sup> 그림 7에 오디오 파일이 링크되어 있음  
(<http://prosody.kaist.ac.kr/KLIP2007/tts2pcm.wav>)

합성결과의 품질을 음성합성 시스템에서 보장할 수 없기 때문에, 앞으로 다양한 형태의 후처리 방안에 대해서 연구가 진행되어야 할 것으로 생각된다.

본 논문에서는 단음절 어휘를 대상으로 담화표지의 음운특징을 분석하였는데 보다 자연스럽게 인식성이 좋은 음성합성 시스템을 위해서는 다양한 형태의 담화표지에 대한 음성언어적 분석과 연구가 필요하다.

### Acknowledgements

본 연구는 뇌과학연구센터와 21세기 프론티어 연구개발사업(인간기능 생활지원 지능로봇 기술개발사업)을 통하여 산업자원부의 지원을 받았음

### 참고문헌

- [1] 임규홍, 국어 담화표지 '인자'에 대한 연구, 담화와 인지 2, 1996.
- [2] 이정애, 국어 화용표지의 연구, 월인, 2002.
- [3] Andrei Popescu-Belis, Sandrine Zufferey, Automatic Identification of Discourse Markers in Multiparty Dialogues, ISSCO Working Paper 65, 2006.
- [4] 임규홍, 성별에 따른 국어 담화 표지 사용 모습, 한국어문학회, 제 83 호, pages 93-113, 2004.
- [5] 최전승 외, 전북방언의 특징과 변화의 방향, 어학 19, pages 49-144, 1992.
- [6] 김하수, 언어 행위와 듣는 이의 신호에 관한 화용론적 분석 시도, 말 14, pages 55-70, 1989.
- [7] 안주호, 한국어 담화표지 분석, 말 17, 1992.
- [8] 김태엽, 국어 담화표지의 유형과 담화표지되기, 우리말글 19, pages 1-23, 2000.
- [9] 김태엽, 담화표지되기과 문법화, 우리말글 26, pages 61-80, 2002.
- [10] 구종남, 담화표지 '뭐'의 문법화와 담화 기능, 국어문학 35, pages 5-32, 2000.
- [11] 이해영, 담화표지 '글쎄'의 담화기능과 사용의미, 이화어문논집 13, pages 129-150, 1994.
- [12] Peter A. Heeman, James F. Allen, Intonational Boundaries, Speech Repairs and Discourse Markers: Modeling Spoken Dialog, In proceedings of the 35th annual meeting on Association for Computational Linguistics, pages 254-261, 1997.
- [13] 김영철, 우리말 담화표지의 기능 고찰, 한국언어문 학 제 52집, pages 23-36, 2004.